

# Semantic Web, SKOS und Linked Data

Kai Eckert

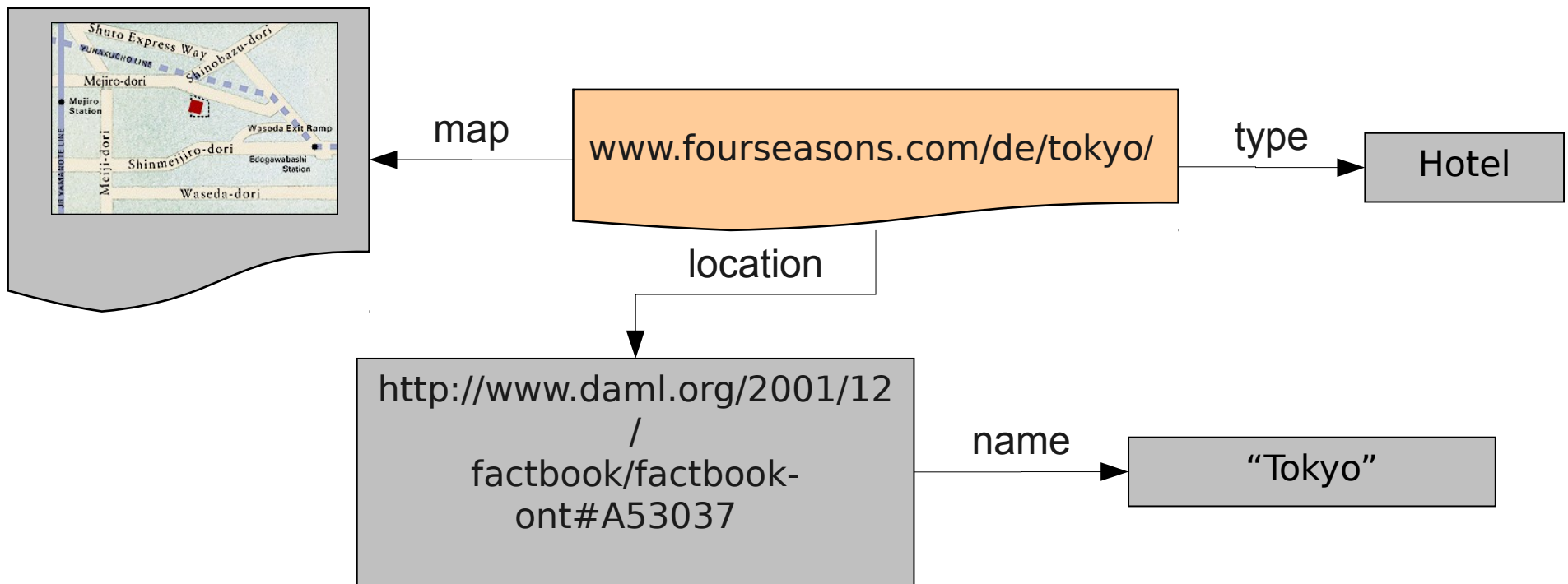
4. Leipziger Kongress für Information und Bibliothek, 16. März 2010

Workshop in Kooperation mit dem Kompetenzzentrum Interoperable Metadaten (KIM)

# Was brauchen wir?

- Eine **standardisierte Syntax**
  - Damit Metadaten überhaupt als solche erkannt werden!
  - W3C: XML, **RDF**, **RDF Schema**, **OWL**
- Ein oder mehrere **standardisierte Vokabularien**
  - Damit Suchmaschinen, Produzenten und Verbraucher die gleiche Sprache sprechen
- Jede Menge Ressourcen, die durch **Metadaten** beschrieben sind.

# RDF: Resource Description Framework



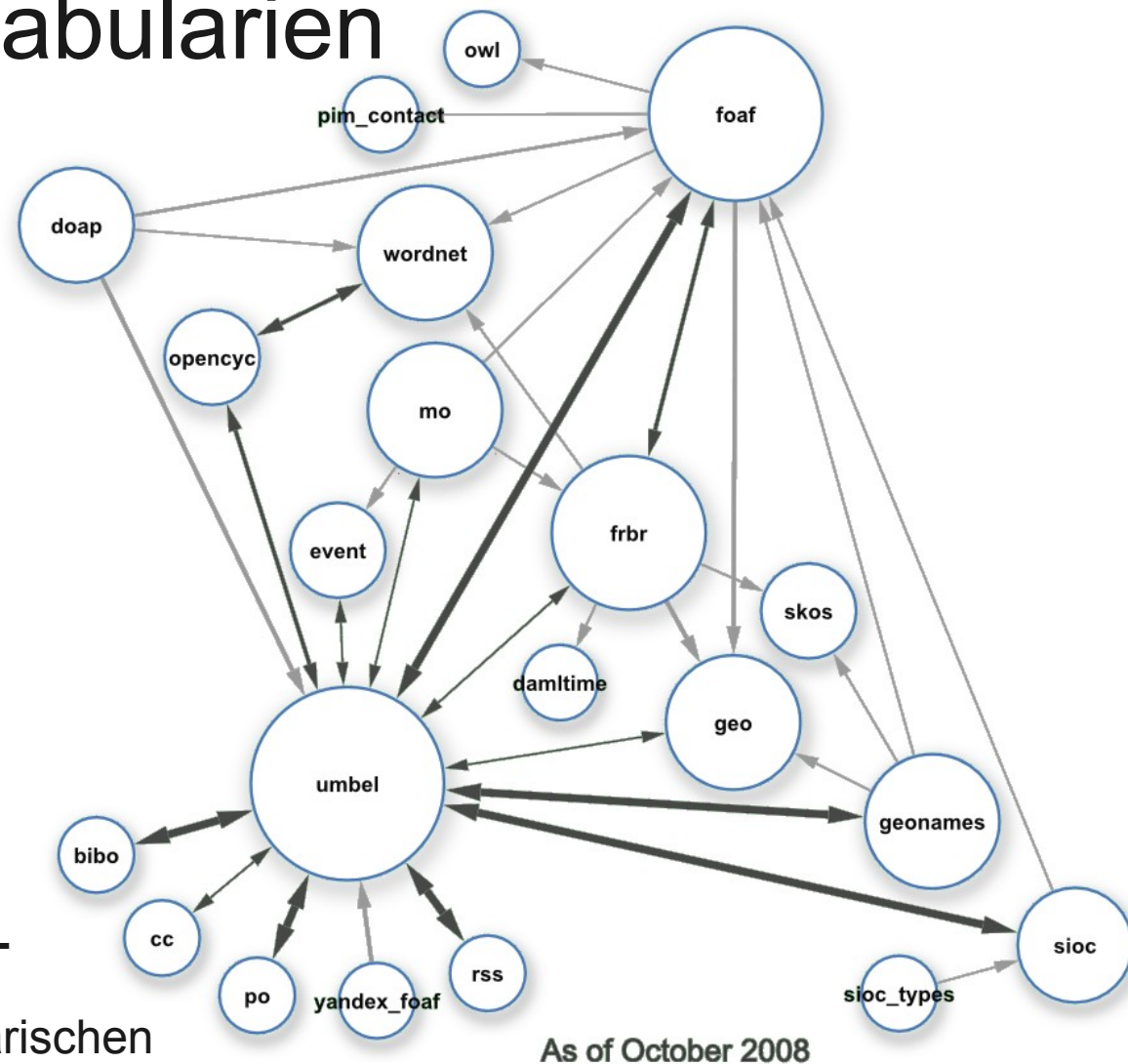
# Vokabularien

- Dublin Core
- FRBR
- FOAF
- SKOS
- SIOC

Das Vokabular für die RDF-Daten.

Beschreibung in **RDF Schema** und **OWL**

Nicht zu verwechseln mit den bibliothekarischen Vokabularien: **Wertevokabularien**

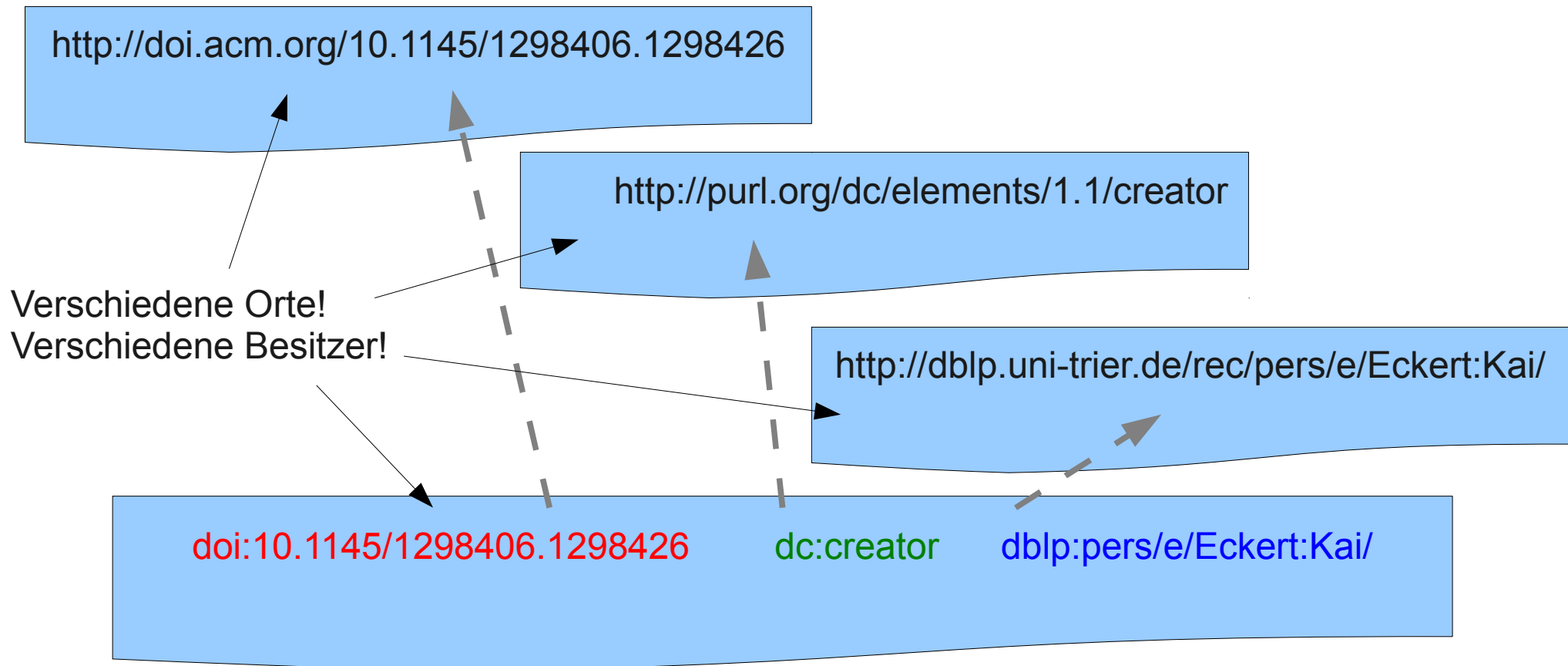


# Was sind RDF-Daten?

- Tripel, die Aussagen (Statements) darstellen:
  - **Subjekt**, **Prädikat**, **Objekt**
  - Beispiel: **Deutschland** **hat als Hauptstadt** **Berlin**.
- Das Subjekt und das Prädikat einer Aussage müssen innerhalb von RDF als URI eindeutig bezeichnet werden.
- Als Objekt können URIs oder Datenwerte (Literele) verwendet werden.
- Beispiel in RDF:
  - **ex:locations/germany** **ex:vocab/capital** **ex:locations/berlin**
  - **ex:locations/germany** **ex:vocab/name** „**Deutschland**“
  - **ex:locations/berlin** **ex:vocab/name** „**Berlin**“

# Die Schönheit von URIs

- Aussagen über Ressourcen können unabhängig von den Ressourcen selbst getätigt werden:



# Die Eindeutigkeit von URIs

- Eine URI beschreibt eindeutig und (hoffentlich) dauerhaft eine ganz bestimmte Ressource.
  - <http://dblp.uni-trier.de/rec/pers/e/Eckert:Kai/>
- Aber: Eine Ressource kann durch verschiedene URIs beschrieben werden
  - <http://dblp.uni-trier.de/rec/pers/e/Eckert:Kai/>
  - [http://ki.informatik.uni-mannheim.de/people/kai\\_eckert.html](http://ki.informatik.uni-mannheim.de/people/kai_eckert.html)
  - <http://de.linkedin.com/in/kaieckert>
  - ...
- **Keine Eineindeutigkeit!**

# Ist das ein Problem?

- Die Äquivalenz von URIs lässt sich jederzeit und von jedem als RDF Statement ausdrücken:
  - <http://de.linkedin.com/in/kaieckert> owl:sameAs <http://dblp.uni-trier.de/rec/pers/e/Eckert:Kai/> ↵
- Dadurch lassen sich insbesondere auch Prädikate (Vokabular) gleichsetzen.
- So lassen sich transparent Datenbestände mit anderen Datenbeständen verknüpfen.
- → **Linked Data**

# Exkurs: URI Syntax

- `<schema> : <Schema-spezifischer Teil>`
- **Beispiele**
  - <http://www.bib.uni-mannheim.de/>
  - <mailto:eckert@bib.uni-mannheim.de>
  - <http://purl.org/dc/elements/1.1/creator>
  - <http://dublincore.org/2008/01/14/dcelements.rdf#creator>
  - <info:lc/vocabulary/countries/af>
  - <urn:nbn:de:gbv:089-3321752945>
  - <urn:isbn:0345391802>
  - <http://doi.acm.org/10.1145/1298406.1298426>
  - <http://dx.doi.org/10.1145/1298406.129842>

# Exkurs: Verschiedene Schemata

- http://...
  - Lediglich eine registrierte Domain notwendig, um Eindeutigkeit zu sichern.
  - Möglichkeit, URIs mit beliebigen Inhalten zu dereferenzieren (RDF!)
- PURL
  - Dienst von OCLC mit dem Vorteil, dass URIs dereferenzierbar bleiben, selbst wenn man die eigene Domain wieder aufgibt.
- URN
  - Eingeschränkt dereferenzierbar über Resolver-Dienste
  - Aufwändige Registrierung über IANA
- info:
  - Nicht referenzierbar, aber leichter zu registrieren als URN Namespaces.

# Exkurs: Referenzen zu URIs

- URI Schemes
  - <http://www.iana.org/assignments/uri-schemes.html>
- Info URIs (`info:<NID>/<NSS>`)
  - [http://info-uri.info/registry/OAIHandler?verb=ListRecords&metadataPrefix=oai\\_dc](http://info-uri.info/registry/OAIHandler?verb=ListRecords&metadataPrefix=oai_dc)
  - `info:lc`
    - <http://info-uri.info/registry/OAIHandler?verb=GetRecord&metadataPrefix=reg&identifier=info:lc/>
- Uniform Resource Names (`urn:<NID>:<NSS>`)
  - <http://www.iana.org/assignments/urn-namespaces/>
  -

# SKOS: Ein Vokabular für (Werte-) Vokabulare

- SKOS = Simple Knowledge Organization System
- Knowledge Organisation Systems (KOS) bezeichnen Systeme zur Wissensorganisation, wie
  - Normdateien, Glossare, Wörterbücher, Ortsverzeichnisse, Schlagwortkataloge, Klassifikationen, Thesauri, Semantische Netzwerke, Ontologien, ...
- „Simple“ drückt aus, dass auf einige Merkmale zur Repräsentation von KOS verzichtet wird.
- Ein „gemeinsamer Nenner“ zur Repräsentation von Thesauri, Klassifikationen und Schlagwortkatalogen.

# SKOS auf einer Folie

- Zentrale Begriff: Das **Konzept** (`concept`)
- Konzepte werden **zusammengefasst** zu einem `concept-scheme`
- Konzepte können mit **Bezeichnern** in verschiedenen **Sprachen** versehen werden: `label`
- **Identifikation** und **Dokumentation**: `notation`, `note`, `changeNote`, `definition`, `editorialNote`, `example`, `historyNote`, `scopeNote`
- Konzepte können **verlinkt** werden: `broader`, `narrower`, `related`
- Schließlich: **Mapping** und **Gruppierung**

# SKOS Vokabular

- **Concept**
- **ConceptScheme**
- *inScheme*
- *hasTopConcept*
- *topConceptOf*
- *altLabel*
- *hiddenLabel*
- **prefLabel**
- **notation**
- RelatedMatch
- **changeNote**
- **definition**
- **editorialNote**
- **example**
- **historyNote**
- **note**
- **scopeNote**
- **Broader**
- broaderTransitive
- **narrower**
- NarrowerTransitive
- **Related**
- *SemanticRelation*
- Collection
- OrderedCollection
- Member
- MemberList
- broadMatch
- closeMatch
- exactMatch
- *mappingRelation*
- narrowMatch
- relatedMatch

# SKOS XL

- SKOS Extension for Labels (XL)
- Hauptpunkte:
  - Lexikalisch identische Labels müssen nicht identisch sein
  - Erlaubt semantische Relationen zwischen Labels
- Insbesondere zur Implementierung von Relationen wie *acronym*, *use*, *used for*, ...

# Beispiel TheSoz

- Thesaurus der Sozialwissenschaften
- 11.600 Terme, davon 7.750 Deskriptoren (Schlagwörter)
- Komplexe Relationen wie „Use for combination“ oder „Used for combination“.
- Besonderheit: 200 Nicht-Deskriptoren zur Disambiguierung:
  - Beispiel: Der alternative Nicht-Deskriptor „Erhebung“ enthält die Relationen „Use Datengewinnung“, „Use Revolution“ sowie „Use Widerstand“, die alle gleichwertig zu behandeln sind, und behandelt dadurch die Mehrdeutigkeit des Begriffes „Erhebung“.
- Zusätzlich existiert eine parallele Klassifikationshierarchie, der die Deskriptoren zugeordnet werden.

# Beispielimplementierung: TheSoz in SKOS (benutzt SKOS-XL)

Element	Beschreibung / Funktion	SKOS Klasse / Property
DD	Deskriptor	skosxl:prefLabel
ND	Nicht-Deskriptor	skosxl:altLabel
AD	Alternativer Nicht-Deskriptor	skosxl:altLabel
NT	Unterbegriff (Narrower Term)	skos:narrower
BT	Oberbegriff (Broader Term)	skos:broader
RT	Verwandter Term (Related Term)	skos:related
USE	Benutze Y statt X (Use)	<b>ext:USE als Property der Klasse</b> <b>ext:equivalenceRelationship</b>
UF	Gegenteil von USE: Y benutzt statt X (Used For)	ext:UF als Property der Klasse ext:equivalenceRelationship
USK	Benutze Kombination X UND Y (Use Combination)	ext:preferredTermComponent als Property der Klasse ext:CompoundEquivalence
UFK	Gegenteil von USK: Benutzt für X mit Y (Used For Combination)	ext:compoundNonPreferredTerm und ext:preferredTermComponent als Properties der Klasse ext:CompoundEquivalence
translation	Engl. Übersetzung des Terms	<b>ext:hasTranslation</b>
scope	Scope Notes	skos:scopeNote
Notation code	Numerischer Code der systematischen Klassifikation, der der Term zuzuordnen ist	skos:notation

# Beispiel STW

- Standard Thesaurus Wirtschaft, <http://zbw.eu/stw/>
- 5.800 Schlagwörter
- 17.000 Terme
- Spezialität: Übergeordnete hierarchische Klassifikation zur Untergliederung der Schlagwörter mit 500 Klassen
- Ziel: Einheitliche Hierarchie, aber weiterhin Unterscheidung zwischen der Klassifikation und den eigentlichen, zur Verschlagwortung verwendeten, Schlagwörtern.
- Möglich durch Einführung neuer RDF-Klassen, die als Unterklassen von `skos:concept` definiert sind.

# Implementierung STW in SKOS

- Neue Unterklassen von `skos:Concept`:
  - `zbwext:Descriptor`: „Normale“ Konzepte
  - `zbwext:Thsys`: Klassen aus der übergeordneten Hierarchie zur Klassifikation der Konzepte, werden nicht zur Beschreibung von Dokumenten genutzt.
- Neue Unterklasse von `skos:note`:
  - `zbwext:useInsteadNote`: Enthält Hinweise über „Use Instead“-Relationen zwischen Termen, die nicht verloren gehen sollen.

# Weitere SKOS Projekte

- **Standard Thesaurus Wirtschaft (ZBW)**
- **Thesaurus für die Sozialwissenschaften (Gesis)**
- General Multilingual Environmental Thesaurus
- UK Archival Thesaurus
- UK Government Category List
- AGROVOC FAO Agricultural Thesaurus
- WordNet
- LCSH
-



# Die Daten kommen...

15.03.2010 13:51

 [« Vorige](#) | [Nächste »](#)

## Open Access: Bibliotheken geben Katalogdaten frei

 [vorlesen](#) / [MP3-Download](#)

Zum Start des Leipziger [Kongresses für Information und Bibliothek](#) am Montag haben mehrere deutsche Bibliotheken ihre Kataloge zur freien Nutzung ins Internet gestellt. Mit dieser Initiative wollen die Bibliotheken einen Beitrag zum freien Zugriff auf Inhalte leisten und setzen ein Zeichen gegen den kommerziell geprägten Austausch von Metadaten. Über [Linked Open Data](#) sind seit heute mehr als 5,4 Millionen Datensätze frei verfügbar, darunter die kompletten Kataloge der Universitäts- und Stadtbibliothek Köln und des Landesbibliotheksentrums Rheinland-Pfalz. Die Daten stehen unter der [CC0-Lizenz](#), das Verwenden und Weiterverarbeiten der Daten ist damit jedem freigestellt.

Dieser Schritt ist ein Paradigmenwechsel für die deutschen Bibliotheken. "Bibliotheken begrüßen die Open-Access-Bewegung, weil sie sich selbst dem Ziel verpflichtet fühlen, den Zugang zu Wissen möglichst ohne Schranken allen Bürgern

### heise Netze

#### iMonitor – Internet-Störungen

Unser Tool zeigt mit Ihrer Hilfe, an welchen Orten welche Zugänge nicht funktionieren.



### heise Developer

#### jQuery 1.4

JavaScript-Bibliothek drückt aufs Tempo



### iX 3/2010:

#### Force Majeure:

Mehrere Prozessorkerne besser nutzen

# Linked Open Data der Bibliotheken

- Katalog der Universitäts- und Stadtbibliothek Köln (USB)
- Katalog der Hochschulbibliothek der Fachhochschule Köln
- Katalog der Bibliothek/Mediathek der Kunsthochschule für Medien Köln
- Katalog des Landesbibliotheksentrums Rheinland-Pfalz (LBZ), PLB Speyer
- Katalog des Landesbibliotheksentrums Rheinland-Pfalz (LBZ), RLB Koblenz
- Katalog des Landesbibliotheksentrums Rheinland-Pfalz (LBZ), BB Zweibrücken
- → **5,4 Millionen Datensätze**

# Beispiel: New York Times

- Veröffentlicht seit 2009 den New York Times Index als Linked Open Data (Stand: März 2010):
  - 4.978 Personen
  - 3.081 Organisationen
  - 1.910 Orte
- <http://data.nytimes.com/>
- Linked Data:
  - Eigene Daten, Verknüpfungen zu Artikeln, Provenienz-Informationen
  - Verknüpft mit geonames.org, dbpedia, freebase, ...

# Beispiel: DBLP

- DBLP Projekt: <http://www.informatik.uni-trier.de/~ley/db/>
- Veröffentlichung als RDF: <http://www4.wiwiss.fu-berlin.de/dblp/>
- Hauptsächlich Dublin Core und FOAF Vokabular

# Referenzen

- Adolphus (2010): The Semantic Web – a new tool for libraries?
  - [http://info.emeraldinsight.com/librarians/info/viewpoints/semantic\\_web.htm](http://info.emeraldinsight.com/librarians/info/viewpoints/semantic_web.htm)
- Borst, Neubert (2009): Case Study: Publishing STW Thesaurus for Economics as Linked Open Data
  - <http://www.w3.org/2001/sw/sweo/public/UseCases/ZBW/>
- Neubert (2009): Bringing the „Thesaurus for Economics“ on to the Web of Linked Data
  - [http://events.linkeddata.org/ldow2009/papers/ldow2009\\_paper7.pdf](http://events.linkeddata.org/ldow2009/papers/ldow2009_paper7.pdf)
- Zapilko, Sure (2009): Converting the TheSoz to SKOS
  - [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methodenberichte/2009/TechnicalReport\\_09\\_07.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2009/TechnicalReport_09_07.pdf)
- Van Assem et al. (2006): A Method to Convert Thesauri to SKOS
  - <http://thesauri.cs.vu.nl/eswc06/>

# Fragen und Diskussion

- Kontakt: Kai Eckert - [eckert@bib.uni-mannheim.de](mailto:eckert@bib.uni-mannheim.de)